

# 单脉冲搜索在 CRAFTS 巡天数据研究与应用

张彬<sup>1</sup>, 梁楠<sup>1,2</sup>, 于徐红<sup>1</sup>, 陈宗浩<sup>1</sup>, 刘志杰<sup>1</sup>, 游善平<sup>1\*</sup>

(贵州师范大学贵州省信息与计算科学重点实验室, 贵州 贵阳, 550001)<sup>1</sup>

(中国科学院云南天文台, 云南 昆明, 650216)<sup>2</sup>

**摘要:** 对大规模射电天文巡天项目产生的数据进行单脉冲搜索会产生数以亿计的脉冲星候选体。本文处理了 CRAFTS (The Commensal Radio Astronomy FAST Survey) 巡天的 32 万个超宽带数据文件, 得到 5 万个脉冲星候选体。针对 CRAFTS 单脉冲搜索实验结果中脉冲星候选体仅有 1.5% 真正具有脉冲星信号的问题, 提出三个显著区分脉冲星信号与干扰的特征, 并应用到单脉冲筛选流程中。实验表明, 改进后的单脉冲筛选工具将需要人工检查脉冲星候选体的数量降低了 20%, 一定程度上提升了 CRAFTS 巡天数据的单脉冲搜索效率。

**关键词:** 单脉冲搜索; 脉冲星; 色散; CRAFTS; 信噪比; 射频干扰

## 1 引言

脉冲星是一种快速旋转、高磁化的中子星, 被认为是理解超密度物质的关键, 可以用于各种基础物理实验<sup>[2]</sup>。目前, 脉冲星搜索方法主要分为周期性搜索和单脉冲搜索两大类。其中, 周期性搜索通过应用快速傅里叶变换 (FFT) 将时间序列转化到频域, 以识别周期性信号。然后在确定的周期内对原始时间序列数据进行折叠, 以放大周期性信号的信噪比 (SNR)<sup>[3]</sup>。而单脉冲搜索寻找强的, 非周期的脉冲, 并未使用快速傅里叶变换和折叠。单脉冲搜索非常适合发现周期性搜索中无法发现的孤立爆发。它的应用导致了旋转射电暂现源 (Rotating Radio Transients, RRAT)<sup>[4]</sup> 和快速射电暴的 (Fast Radio Bursts, FRB)<sup>[5]</sup> 的发现, 从而激起人们对单脉冲搜索的兴趣。

2003年, Cordes和McLaughlin首次提出了一个理论框架, 运用单脉冲搜索方法来探测射电天文观测数据中的脉冲星信号<sup>[6]</sup>。Deneva在2009年把单脉冲搜索方法应用到阿雷西博 (Pulsar Arecibo L-band Feed Array, PALFA) 七波束观测数据上, 并发现7颗脉冲星<sup>[7]</sup>。同样, Keane在2010年使用单脉冲搜索方法对帕克斯 (Parkes Multibeam Pulsar Survey, PMPS) 多波束数据文件进行重新分析, 发现10颗RRAT<sup>[8]</sup>。2015年, Karako和Argaman在单脉冲理论上, 设计出针对RRAT的搜索工具 RRATtrap, 应用在绿岸 (Green Bank Telescope, GBT) 350MHz数据, 最终探测到18颗RRAT<sup>[9]</sup>。2016年, Devine首次将单脉冲搜索与机器学习结合, 提出RAPID筛选工具, 实现了自动化筛选脉冲星候选体, 并探测到6个潜在的发现<sup>[10]</sup>。

2016年9月位于贵州喀斯特山区的五百米口径球面射电望远镜 (Five-hundred-meter Aperture Spherical radio Telescope, FAST) 进入科学运行阶段, 先后配置了超宽带接收机和L波段19波束接收机设备<sup>[10]</sup>。为充分利用FAST的优异的探测能力, 2018年李菂等人设计并实施了“FAST多科学目标同时扫描巡天” (The Commensal Radio Astronomy FAST Survey, 简称CRAFTS)<sup>[11]</sup>。即同时使用多个数字终端采集脉冲星、中子星、分子谱线、旋转射电暂现源、FRB等多个科学目标的观测数据。鉴于单脉冲搜索方法在各大射电望远镜的成功应用, 本文对CRAFTS超宽带数据文件的单脉冲搜索结果进行研究, 发现CRAFTS单脉冲搜索筛选得到脉冲星候选体存在数以万计的假阳性样本。因此, 我们根据脉冲星信号在色散量与信噪比曲线大致是高斯曲线以及信噪比数据分布是高斯分布的假

基金项目: 国家自然科学基金★(U1631132, U1731238, U1831131) 与贵州省科学技术基金 (黔科合J字 LKS[2010]38号, 黔科合基础-ZK[2021]重点020) 支持。

作者简介: 张彬, 男, 硕士研究生, 研究方向: FAST 数据处理, E-mail: 2379710516@qq.com。

通信作者: 游善平, 男, 博士研究生, 研究方向: FAST 数据处理, E-mail: ysp@gznu.edu.cn。

设，构造三个显著区分脉冲星信号与干扰的特征，旨在缓解出现大量假阳性样本的问题。实验表明添加特征值判断的单脉冲筛选工具与传统方法相比，假阳性样本数量减少了20%。

2 CRAFTS 超宽带数据单脉冲搜索实验

在射电天文观测数据中发现脉冲星通常分为四个阶段：收集、消色散、周期性或单脉冲搜索和人工检查<sup>[2]</sup>。在第一阶段，原始数据以电压时间序列的形式由射电望远镜收集。第二阶段，由于脉冲辐射通过星际介质（Interstellar Medium, ISM）的色散效应，导致较低频率的脉冲比较高频率晚到达<sup>[12]</sup>。消色散是消除这些与频率有关的延迟效应的影响。第三阶段，使用周期性搜索或单脉冲搜索寻找出数据中的脉冲星候选体。第四阶段，对上面判断为脉冲星候选体的数据再进行一次人工检查。

2.1 数据处理

我们先使用 PRESTO（Pulsar Exploration and Search Toolkit）脉冲星搜索工具处理了 CRAFTS2017 年 8 月至 2018 年 5 月的超宽带数据，处理流程包括：去干扰、消色散、单脉冲搜索等步骤；然后，应用单脉冲筛选工具 RRATtrap 从 PRESTO 的 Single\_pulse\_search.py 输出文件中筛选出脉冲星候选体。

单脉冲筛选工具 RRATtrap 根据脉冲星信号如下特性，将脉冲星信号与干扰区分开。1、脉冲星信号出现在一定（Dispersion Measure, DM）范围内。在最佳的 DM 时检测到峰值信噪比，而在该 DM 上下信噪比下降，这是由于脉冲在不正确的 DM 消色散导致脉冲展宽造成的。2、由于信号在最佳 DM 检测到峰值信噪比，预计来自地面的信号（即射频干扰）将在 DM 为  $0\text{ pc cm}^{-3}$  达到峰值。同时射频干扰不会受到色散效应的影响，因此出现在一个非常大的 DM 范围内<sup>[9]</sup>。

图 1 是单脉冲筛选工具 RRATtrap 探测到已知脉冲星 PSR B0540+21 的诊断图。左侧子图纵坐标为试验 DM 范围，横坐标为观测时间，CRAFTS 超宽带数据每个观测文件为 52 秒，右侧子图描述试验 DM 对应的信噪比大小。可以看到呈仿锤状的单脉冲事件组（在相邻的 DM 与时间窗口内所有信噪比大于 5 的事件）在最佳 DM  $78\text{ pc cm}^{-3}$  取得峰值信噪比 68，而在该 DM 上下，信噪比下降。其中，强的脉冲星信号被标记为 Excellent；弱的脉冲星信号被表标记为 Very good；不规则的脉冲星信号被标记为 Good 或 Ok；射频干扰被标记为 RFI；宇宙噪音被标记为 Noise。在图 2 中，我们看到在较宽的 DM 范围上极强的射频干扰的实例。此外，发现在  $t = 20 - 40\text{ s}$ ，极强的射频干扰被错误标记为脉冲星信号。

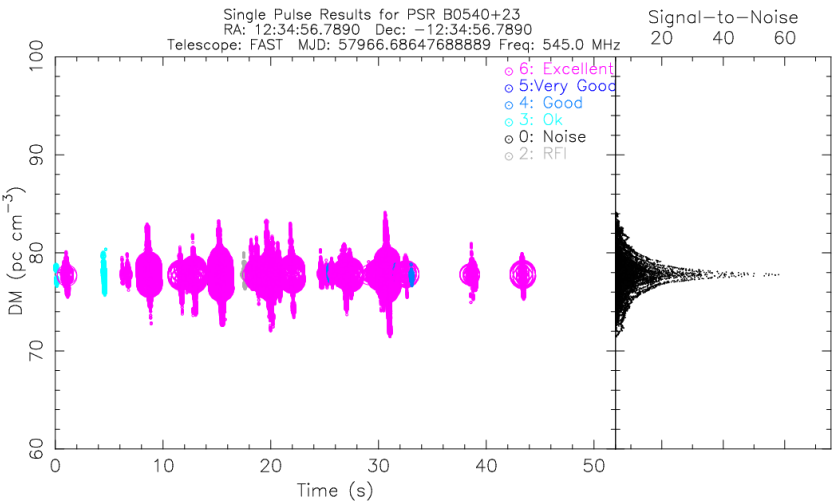


图 1 RRATtrap-探测到脉冲星 PSR B0540+21 诊断图

Fig 1 Diagnostic picture of the pulsar PSR B0540+21 detected by RRATtrap -

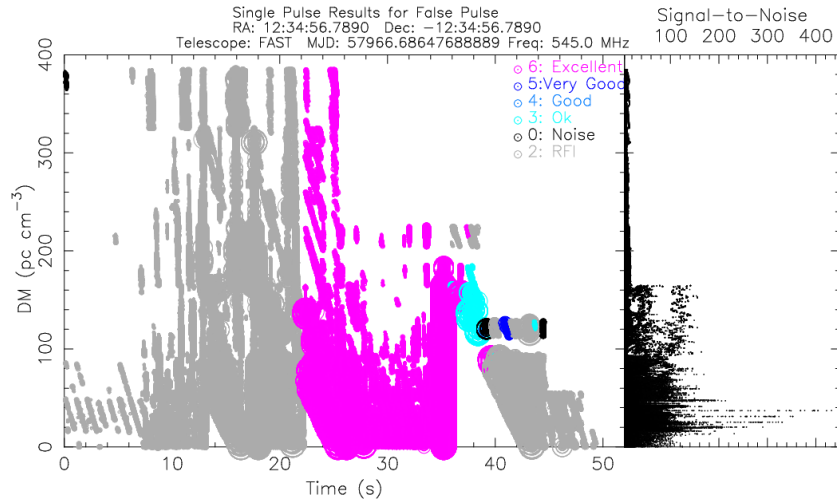


图 2 RRATtrap 筛选出假阳性实例诊断图

Fig 2 RRATtrap screened out false positive case diagnosis diagram

本次实验选取 CRAFTS2017 年 8 月至 2018 年 5 月共计约 32 万个超宽带数据文件<sup>[1]</sup>进行单脉冲搜索实验。由于本次处理的数据规模较大，对所有数据进行全面的人工检查工作量非常大。因此，我们先使用单脉冲筛选工具 RRATtrap 初步筛选出脉冲星候选体。然后，我们对数据文件中标记为 Excellent、Very Good、Good、Ok 共计约 5 万颗脉冲星候选体的诊断图进行手工检查，发现仅有 772 个诊断图真正具有脉冲星信号，对应 101 颗已知脉冲星。通过式 1 计算的假阳性率达到 98.5%，意味着每 100 个脉冲星候选体中，仅有 1 至 2 个包含脉冲星信号（已知或新发现的）。由于 CRAFTS 超宽带数据文件中包含的具体脉冲星数量未知，本文未对假阴性情况进行分析。

$$FPR = \frac{FP}{FP+TP} \quad (1)$$

其中， $FP$  是没有脉冲星信号的数据文件被标记为候选脉冲星的文件数量； $TP$  是含有脉冲星信号的数据文件被标记为候选脉冲星的文件数量。

### 3 区分脉冲星信号与干扰的特征

CRAFTS 超宽带数据单脉冲搜索结果表明：脉冲星候选体仅有 1.5 % 真正具有脉冲星信号，这主要是未能成功区分脉冲星信号与干扰造成的，如图 2 所示。本节首先探讨了不正确的 DM 对信噪比的影响。然后，构造了三个显著区分脉冲星信号与干扰的特征。并在探测到的 101 颗脉冲星样本中选取同时具有脉冲星信号、射频干扰和宇宙噪音的 79 颗脉冲星作为样本，计算它们在三个特征取值情况。最后，根据脉冲星信号、射频干扰和宇宙噪音在三个特征取值分布的差异，提出合理阈值应用在单脉冲筛选工具上，对脉冲星候选体进行进一步筛选。

#### 3.1 DM 与信噪比的曲线

Cordes 与 McLaughlin 探讨了不正确的 DM 对单个脉冲信噪比的影响<sup>[6]</sup>。通常，对一个特定的 DM 的时间序列进行多次下采样并重新搜索，当有效的采样时间最接近脉冲宽度时，脉冲达到最高的信噪比。在不同的 DM 信道中，随着试验 DM 与真实 DM 的偏差越大，信噪比越小，如图 1 所示。测量的信噪比  $S(\delta DM)$  与真正的信噪比  $S$  比值与 DM 的偏差  $\delta DM$  满足如下关系<sup>[5]</sup>：

$$\frac{S(\delta DM)}{S} = \frac{\sqrt{\pi}}{2} \zeta^{-1} \text{erf} \zeta \quad (2)$$

这里:

$$\zeta = 6.91 \times 10^{-3} \delta DM \frac{\Delta \nu_{MHz}}{W_{ms} \nu_{GHz}^3} \quad (3)$$

其中,  $\text{erf}$  是误差函数;  $\delta DM$  是实验 DM 与真实 DM 偏差;  $\Delta \nu$  是总带宽;  $\nu$  是中心频率;  $W$  是脉冲宽度, 单位为毫秒。

利用这两个式(2)(3), 给出天体物理脉冲的信噪比和宽度, 我们可以计算出预期的 DM 偏差  $\delta DM$ 。在图 3 中, 展示了在中心频率  $546 MHz$ , 时间分辨率为  $100 us$  和匹配滤波使用 30 的 PSR B2000+40 一组单脉冲事件预期和拟合信噪比下降曲线。在这两种情况下, 信噪比峰值处的 DM 被认为是真正的 DM。我们计算预期的信噪比下降, 利用峰值信噪比和匹配滤波得到脉冲宽度, 这与实际脉冲宽度很接近。另一方面, 为了计算拟合的信噪比下降, 首先, 使用非线性最小二乘法 (NLS) 对原始数据进行回归得到拟合的峰值, 然后, 使用相同的方程计算拟合的信噪比下降。图 3 可以看出, 预期和拟合的 DM 与信噪比曲线大致是高斯曲线。使用观测到的峰值信噪比和匹配滤波得到脉冲宽度, 预期的信噪比下降与拟合值在峰值右侧相当接近, 但在左侧出现偏差。这是匹配滤波得到的脉冲宽度与实际脉冲宽度的偏差造成的。

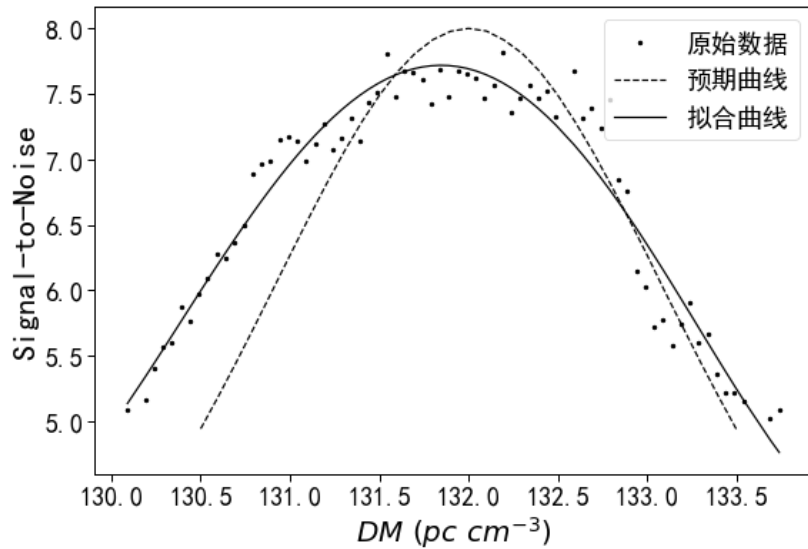


图 3 PSR B2000+40 一组单脉冲事件色散与信噪比曲线图

Fig 3 PSR B2000+40 a set of single pulse event dispersion and signal-to-noise ratio curve

### 3.2 三个显著特征

Lorimer 指出脉冲星的轮廓可以简化为高斯型 (对大多数脉冲星来说是一个合理的近似) [14]。按照 Cordes 与 McLaughlin 提出单脉冲搜索理论, 天体物理脉冲的 DM 和信噪比曲线大致是高斯曲线<sup>[5]</sup>, 如图 3 所示。由于高斯曲线通常是对称, Di Pang 提出了两种对称特征,  $SI_{DM}$  和  $SI_{S/N}$ , 来表征单脉冲事件组的 DM 与信噪比曲线的对称性<sup>[15]</sup>。  $SI_{DM}$  与  $SI_{S/N}$  的计算方法如下:

$$SI_{DM} = \frac{\min(\max DM - peakDM, peakDM - \min DM)}{\max(\max DM - peakDM, peakDM - \min DM)} \quad (4)$$

这里,  $peakDM$  是信噪比峰值对应的  $DM$ ;  $\max DM$  是单脉冲事件组最大的  $DM$ ;  $\min DM$  是单脉冲事件组最小的  $DM$ 。

$$SI_{S/N} = \frac{\min(\sum S/N_{left}, \sum S/N_{right})}{\max(\sum S/N_{left}, \sum S/N_{right})} \quad (5)$$

这里,  $\sum S/N_{left}$  是信噪比峰值左侧所有单脉冲事件的信噪比求和;  $\sum S/N_{right}$  是信噪比峰值右侧所有单脉冲事件的信噪比求和。

由式 (4) 和 (5) 定义的  $SI_{DM}$  和  $SI_{S/N}$  的取值范围都在 0 到 1, 这两个特征取值越高, 则越对称, 相反不遵循方程 (2) 描述规律的射频干扰, 通常在  $DM$  与信噪比空间是单调递减 (或递增)。所以, 它们通常具有接近于 0 的对称值, 可以与脉冲星信号进行区分。如图 4 和 5 所示, 分别展示了脉冲星样本的  $SI_{DM}$  和  $SI_{S/N}$  对称值, 并对脉冲星信号, 射频干扰和噪音进行了对比。可以看出: 脉冲星信号通常具有较大的对称值, 而射频干扰和噪音的对称值相对较小, 这表明脉冲星信号比干扰更对称。此外, 发现一些噪音具有大的对称值, 这些通常是高斯噪声。

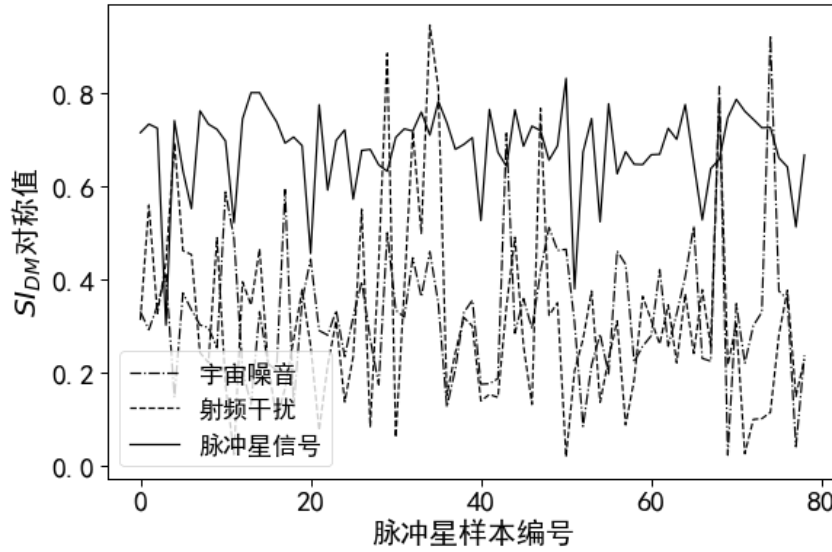
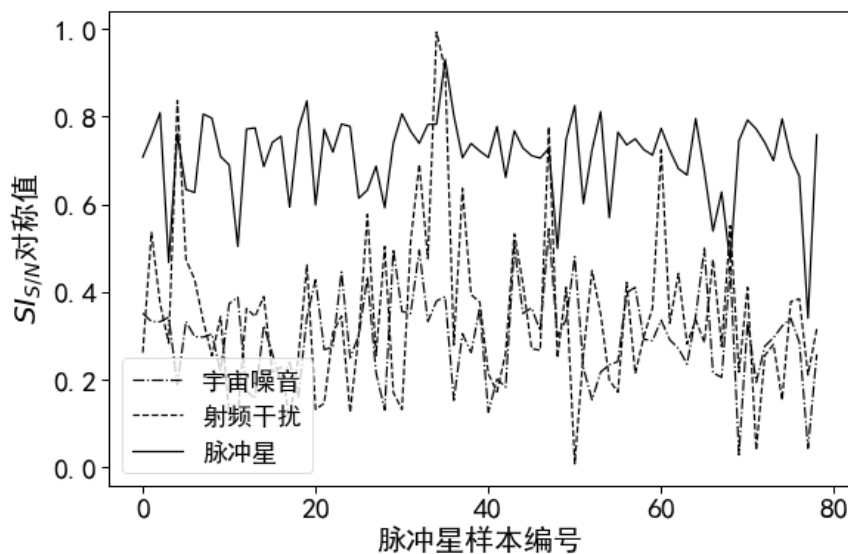


图 4 脉冲星样本  $SI_{DM}$  对称

Fig 4  $SI_{DM}$  Symmetry values of pulsar samples

图 5 脉冲星样本  $SI_{S/N}$  对称值Fig 5  $SI_{S/N}$  Symmetry values of pulsar samples

Cordes 与 McLaughlin 预测在没有任何宇宙噪音和射频干扰的情况下, 信噪比的数据分布是高斯分布<sup>[6]</sup>。峰度是描述总体数据分布与高斯分布的陡缓程度统计量。为了判断脉冲星样本信噪比的数据分布是否符合高斯分布, 我们使用 kurtosis 峰度值来表征单脉冲事件组内信噪比的数据分布与高斯分布的符合程度。kurtosis 峰度值的计算方法如下:

$$kurtosis[S/N] = E\left[\left(\frac{S/N - \mu}{\sigma}\right)^4\right] = \frac{E[(S/N - \mu)^4]}{(E[(S/N - \mu)^2])^2} \quad (6)$$

这里,  $S/N$  是单脉冲事件的信噪比;  $\mu$  是单脉冲事件组内平均信噪比;  $\sigma$  是单脉冲事件组内信噪比的方差。

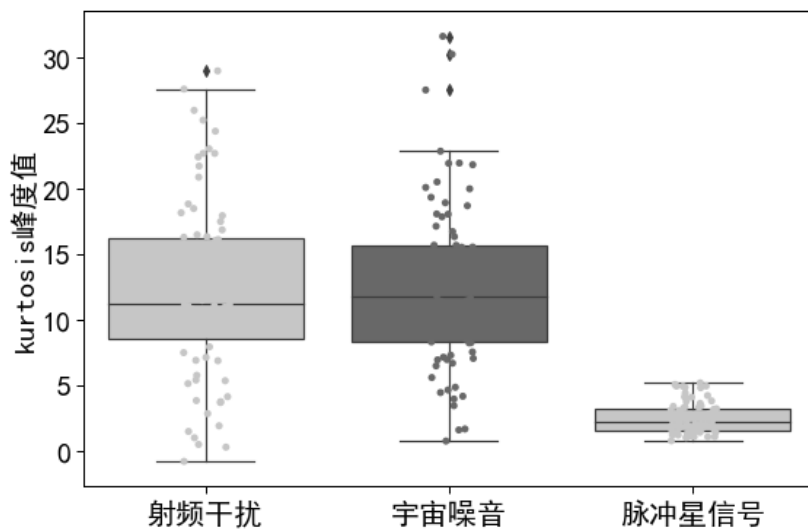


图 6 脉冲星样本 kurtosis 峰度值

Fig 6 kurtosis value of pulsar sample

当式 (6) 定义的 kurtosis 峰度值是 3 时, 信噪比的数据分布服从高斯分布, 随着 kurtosis 峰度值与 3 的差距越大, 其分布形态的陡缓程度与高斯分布的差异程度越大。如图 6 所示, 用箱线图展示了脉冲星样本的 kurtosis 峰度值分布情况, 并对脉冲星信号, 射频干扰和噪音进行了对比。可以看出: 脉冲星信号峰度值集中于 3 附近, 而射频干扰和噪音的峰度值分散, 且与 3 差距较大。表明脉冲星信号信噪比数据分布近似服从高斯分布, 而大部分干扰与噪音不具有此规律。

图 4、5、6 可以明显看出脉冲星信号、射频干扰和宇宙噪音在三个特征分布的差异。因此, 我们根据脉冲星信号与干扰特征值的分布区间, 如表 1 所示, 选择合适的特征阈值, 剔除不满足阈值的脉冲星候选体, 从而达到降低假阳性率的目的。(比如: 设置脉冲星信号特征  $SI_{DM}$  最小值 0.1 作为阈值, 对低于阈值的候选体视为干扰处理。那么, 特征值在 0.02 到 0.1 区间的射频干扰将不会被错误的标记为脉冲星候选体。) 同时, 我们选取不同阈值进行实验, 并统计其对应的假阳性率和遗漏脉冲星信号的结果, 如图 7 所示。可以看出: 随着阈值变大或者阈值区间缩小, 假阳性率不断减小; 与此同时, 遗漏发现的脉冲星数量在不断增加。而我们做脉冲星搜寻工作, 首要前提是保证观测数据中所有脉冲星信号不被遗漏。因此, 我们选取表 1 中脉冲星信号特征  $SI_{DM}$  最小值 0.1,  $SI_{S/N}$  最小值 0.29 作为阈值; 特征 kurtosis 的区间-1.54~27.11 作为阈值区间, 对低于阈值或不在阈值区间的脉冲星候选体视为射频干扰或宇宙噪音处理。

表 1 三个特征在三种类别的分布区间

Table 1 Value ranges of the three characteristics in the three categories

特征	脉冲星信号		射频干扰	宇宙噪音
	CRAFTS	PMPS	CRAFTS	CRAFTS
$SI_{DM}$	<b>0.10~0.97</b>	0.25~0.90	0.02~0.92	0.01~0.94
$SI_{S/N}$	<b>0.29~0.96</b>	0.26~0.96	0.01~0.99	0.02~0.54
kurtosis	<b>-1.54~27.11</b>	-0.81~23.38	-0.80~28.95	0.74~31.58

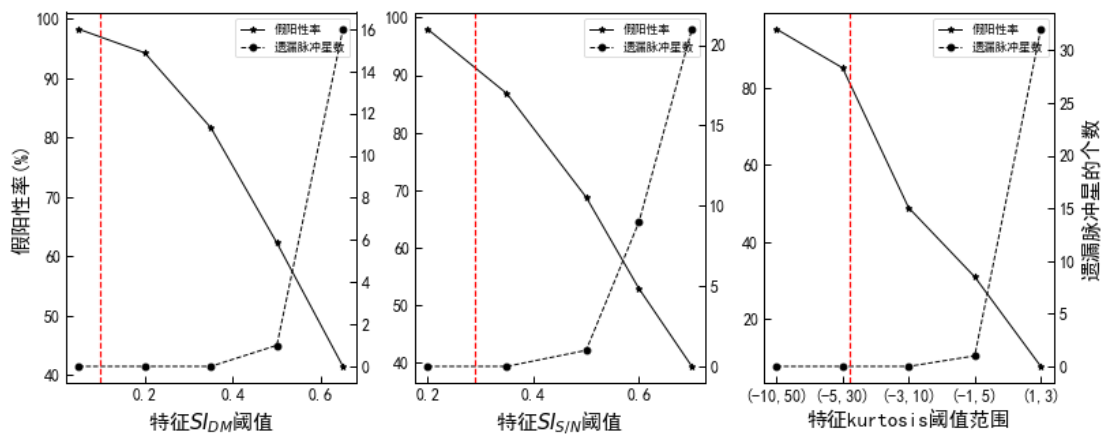


图 7 各个特征在不同阈值假阳性和遗漏脉冲星的情况, 其中红色竖线代表所选择的阈值  
Fig 7 Each feature has different thresholds for false positives and missing pulsars, where the red vertical line represents the selected threshold

此外, 为了证实我们所述特征是否在其它观测数据有效, 我们对帕克斯多波束数据进行了实验。根据张松波公开的帕克斯单脉冲数据库<sup>[16]</sup>, 构建了一个 PMPS (Parkes Multibeam Pulsar

Survey) 数据集, 其中包含帕克斯发现的部分脉冲星信号。我们计算得到 **PMPS** 数据集中脉冲星信号在三个特征值的分布情况, 并与 **CRAFTS** 进行对比, 如表 1 所示。可以发现: 它们的特征值分布十分接近, 尽管它们在消色散网格、周围射频干扰的环境等存在显著差异。因此, 我们所述的特征阈值同样能够将帕克斯数据中脉冲星信号从干扰中区分开。

我们将上述阈值和阈值区间应用到单脉冲筛选工具上, 对脉冲星候选体进行重新筛选。实验结果如表2所示: 可以看出添加单个特征阈值判断的单脉冲筛选工具能在一定程度上降低假阳性率; 并且三个特征阈值组合判断假阳性率从98.5%降低到78.4%。表明添加特征阈值判断的单脉冲筛选工具提升了脉冲星搜索效率。

表2 添加特征后假阳性率结果  
Table 2 False positive rate results after adding features

特征	假阳性率 (FPR)
$SI_{DM}$	97.9%
$SI_{S/N}$	93.8%
kurtosis	82.5%
$SI_{DM} + SI_{S/N} + \text{kurtosis}$	<b>78.4%</b>

4 总结与展望

本文使用**PRESTO**的单脉冲搜索方法对**CRAFTS**超宽带数据文件进行的实验表明, 现有的单脉冲筛选工具难以区分真实的信号与噪音或射频干扰返回的信号, 造成数以万计的假阳性样本出现, 使得脉冲星的发现变得困难。因此, 我们先提出三个显著区分脉冲星信号与干扰的特征。并选取同时具有脉冲星信号、射频干扰和宇宙噪音的79颗脉冲星样本, 计算它们在三个特征取值情况。然后, 根据脉冲星信号、射频干扰和宇宙噪音在三个特征取值分布的差异, 提出合理阈值应用在单脉冲筛选工具上, 用于对脉冲星候选体进行严格的判断。最后, 使用添加特征阈值判断的单脉冲筛选工具对脉冲星候选体进行重新实验。结果表明: 在保证所有脉冲星信号都不被遗漏的情况下, 假阳性率从98.5%降低到78.4%。因此, 本文所述三个特征在实验的验证下证明了具有实用性和有效性, 可以助于单脉冲搜索在**CRAFTS**巡天数据的应用。

**致谢:** 本工作在 **FAST** (500 米口径球面射电望远镜) 数据基础上完成。**FAST** 是由中国科学院国家天文台运行和管理的国家大科学装置。我们感谢中国科学院天文大科学研究中心 **FAST** 重大成果培育项目对本工作的资助。

References (参考文献)

[1] Li D, Wang P, Qian L, et al. FAST in space: Considerations for a Multi-beam multipurpose Survey using china' s 500-m aperture spherical radio telescope (FAST) [J]. IEEE Microwave Magazine, 2018, 19(3):112-119.

[2] Lorimer D R, Kramer M. Handbook of Pulsar Astronomy[M].Cambridge: Cambridge University Press. 2004.

[3] Larsson S. Parameter estimation in epoch folding analysis[J]. Astronomy and Astrophysics Supplement Series, 1996, 117(1):197-201.

[4] McLaughlin M A, Lyne A, Lorimer D, et al. Transient radio bursts from rotating neutron stars[J].Nature, 2006, 439, 817 - 820.

- [5] Lorimer D R, Bailes M, McLaughlin M A, et al. A bright millisecond radio burst of extragalactic origin[J]. *Science*, 2007, 318, 777 - 780.
- [6] Cordes J M, McLaughlin M A. Searches for Fast Radio Transients[J]. *The Astrophysical Journal*, 2003, 596(2):1142-1154.
- [7] Deneva J S, Cordes J M, McLaughlin M A, et al. Arecibo Pulsar Survey Using ALFA: Probing Radio Pulsar Intermittency and Transients[J]. *The Astrophysical Journal*, 2009, 703(2):2259-2274.
- [8] Keane E F, Ludovici D A, et al. Further searches for Rotating Radio Transients in the Parkes Multi-beam Pulsar Survey[J]. *Monthly Notices of the Royal Astronomical Society*, 2010, 401(1):1057-1068.
- [9] Karako-Argaman C, et al. Discovery and Follow-up of Rotating Radio Transients with the Green Bank and LOFAR Telescopes[J]. *The Astrophysical Journal*, 2015, 809(2):67-85.
- [10] Devine T R, Katerina G P, Maura M L. Detection of Dispersed Radio Pulses: A machine learning approach to candidate identification and classification[J]. *Monthly Notices of the Royal Astronomical Society*, 2016, 459(1):1519 - 1532.
- [11] 刘鹏, 王培, 李葭, 等. FAST 19 波束脉冲星漂移扫描巡天模拟[J]. *天文学进展*, 2018, 36(2): 173-188.
- [12] 南仁东, 李会贤. FAST的进展——科学、技术与设备[J]. *中国科学: 物理学 力学 天文学*, 2014, 44(10): 1063-1074.
- [13] James M Cordes. Pulsar observations-Propagation Effects, searching. Distance Estimates, Scintillations and VLBI [J]. *The Astronomical Society of the Pacific*. 2002, 278-286.
- [14] Lorimer D R, Faulkner A J, et al. The Parkes Multibeam Pulsar Survey - VI. Discovery and timing of 142 pulsars and a Galactic population analysis[J]. *Monthly Notices of the Royal Astronomical Society*, 2006, 372(2):777-800.
- [15] Di Pang, Katerina Goseva-Popstojanova, et al. A novel single-pulse search approach to detection of dispersed radiopulses using clustering and supervised machine learning[J]. *Monthly Notices of the Royal Astronomical Society*, 2018, 459(2):1519-1541.
- [16] Zhang S B, Hobbs G, et al. Parkes transient events: I. Database of single pulses, initial results and missing FRBs[J]. *The Astrophysical Journal*, 2020, 888(1):62-74.

## Research and Application of Single Pulse Search in CRAFTS Sky Survey Data

Zhang Bin<sup>1</sup>, Liang Nan<sup>1,2</sup>, Yu Xuhong<sup>1</sup>, Chen Zonghao<sup>1</sup>, Liu Zhijie, You Shanping<sup>1\*</sup>

(Guizhou Provincial Key Laboratory of Information and Computing Science, Guizhou Normal University, Guiyang 550001)<sup>1</sup>

(Yunnan Observatories, Chinese Academy of Sciences, Kunming, Yunnan 650216) <sup>2</sup>

**Abstract:** A single pulse search on the data generated by the large-scale radio astronomy survey project will generate hundreds of millions of pulsar candidates. This paper processed 320,000 ultra-wideband data files of the CRAFTS (The Commensal Radio Astronomy FAST Survey) survey, and obtained 50,000 pulsar candidates. In order to solve the problem that only 1.5% of the pulsar candidates in the results of CRAFTS single pulse search experiment really have pulsar signal, three features that distinguish pulsar signals from interference are proposed and applied to the single pulse sifting process. Experiments show that the improved single pulse sifting method will reduce the number of pulsar candidates that need to be manually checked by 20%, which improves the single pulse search efficiency of CRAFTS survey data to a certain extent.

**Keywords:** single pulse search; pulsar; dispersion; CRAFTS; signal-to-noise ratio; Radio frequency interference